

# Developing Achievement Test: A Research for Assessment of 5th Grade Biology Subject

Nilay Sener<sup>1</sup> & Erol Tas<sup>2</sup>

<sup>1</sup> The Ministry of National Education Secondary School Özcan Duran Karagöl, Muğla, Turkey

<sup>2</sup> Science Education, Ordu University, Ordu, Turkey

Correspondence: Nilay Sener, The Ministry of National Education Secondary School Özcan Duran Karagöl, Muğla, Turkey. E-mail: nilsener471@gmail.com

Received: December 28, 2016

Accepted: January 20, 2017

Online Published: February 16, 2017

doi:10.5539/jel.v6n2p254

URL: <http://doi.org/10.5539/jel.v6n2p254>

## **Abstract**

The purpose of this study is to prepare a multiple-choice achievement test with high reliability and validity for the “Let’s Solve the Puzzle of Our Body” unit. For this purpose, a multiple choice achievement test consisting of 46 items was applied to 178 fifth grade students in total. As a result of the test and material analysis performed during the test development process, difficulty, distinctiveness, and item-total correlation coefficients of the materials were calculated. For the validity study, a table of specifications was prepared and the Content Validity Index (CVI) was found to be 0.95 by taking an expert opinion. As a result of the analysis, 8 items were removed from the test and the KR-20 reliability coefficient of the final test consisting of 38 items was calculated as 0.87. As a result of the item analyses, while item difficulty indices were valued between 0.30 and 0.74, item distinctiveness indeces were valued between 0.31 and 0.71. The average difficulty of the test was calculated as moderate (0.56) and its distinctiveness was calculated as very good (0.49).

**Keywords:** biology, developing achievement test, item analysis, reliability, science education, validity

## **1. Introduction**

In the researches related to the 5<sup>th</sup> grade biology subjects covering the transition period from primary school to secondary school, it is revealed that the students have deficient level of knowledge and alternative concepts in “*Nutrients and their characteristics*”, “*Digestion of nutrients*” and “*Excretory System in our body*” which are included in “Let’s Solve the Puzzle of Our Body” unit (Banet & Núñez, 1997; Carvalho, Silva, & Clément, 2003; Güngör, 2009; Güngör & Özgür, 2009; Núñez & Banet, 1997; Patrick & Tunnicliffe, 2010). The deficient level of knowledge level of the students from their early ages led the researchers to make studies about eliminating the lack of information and alternative concepts of the students in these subjects.

In educational studies, many measurement tools such as interviews, open-ended questions, concept maps, tests can be used to determine the level of students’ understanding of knowledge and concepts. Qualitative ones from these researches can work with fewer participants, enabling more in-depth research; however, with quantitative ones a large audience can be reached with more participants (Griffard, 2001). Multiple choice tests are very suitable measurement tools for determining the level of knowledge of different subjects of many students at different academic levels (Burton, Sudweeks, Merrill, & Wood, 1991). Multiple-choice tests also enable students to determine the misconceptions they have by including inaccuracies they have in the options (Treagust, 1988).

When the literature is examined, there are many researches carried out with the 6<sup>th</sup> and 7<sup>th</sup> grade (Akgündüz, 2013; Aydede & Matyar, 2009; Erdoğan, 2010; İnel, 2009; Kiras, 2013) or researches carried out with the university students and the teachers, in relation to sub topics and concepts in the content of the “*Let’s Solve the Puzzle of Our Body*” unit (Çardak, 2005; Patrick & Tunnicliffe, 2010; Prokop & Faněovičová, 2006). However, there are very few studies related to Let’s Solve the Puzzle of Our Body unit in 5<sup>th</sup> grade (Güngör & Özgür, 2009). As there are only a few researches related to this subject, it has been needed to develop a reliable and valid mesurement tool for “*Let’s Solve the Puzzle of Our Body*” unit since it is one of the essential requirements especially for science teaching of the 5<sup>th</sup> grade, which is the first step of the secondary school grade.

## 2. Method

### 2.1 Sample

Pilot application of the study was carried out with a total of 178 5<sup>th</sup> grade students, 89 female and 89 male students studying at two central schools in Samsun city of Turkey. The schools of whom have been randomly selected by lot, among the schools belonging to Turkish Ministry of National Education without considering their academic success. The distribution of the sample according to the schools and sex is shown in Table 1.

Table 1. The number of schools and students participated in the pilot study

Gender	Secondary School A		Secondary School B		Total	
	f	%	f	%	f	%
Girl	43	48.3	45	50.6	88	49.4
Boy	46	51.7	44	49.4	90	50.6
<b>Total</b>	<b>89</b>	<b>100</b>	<b>89</b>	<b>100</b>	<b>178</b>	<b>100</b>

### 2.2 Data Collecting Instrument

In this research, Let's Solve the Puzzle of Our Body Unit Achievement Test was used as a collecting data tool. The aim of using multiple-choice testing as an achievement test is to allow the ability to measure many sub-concepts of the unit taught in the research, to make it easy to evaluate and to enable to measure how much it has been learned (Marx et al., 2004).

The achievement test was prepared by taking into consideration the objectives of the "*Let's Solve the Puzzle of Our Body*" unit which is included in the 5<sup>th</sup> grade Ministry of National Education (MoNE) Science Curriculum to be used in the research. Regarding the level of readiness of 5<sup>th</sup> grade students, 46 multiple choice test items with four options were created. While creating the test items, the questions of the achievement test were created by the researcher by examining the 5<sup>th</sup> grade Science course books prepared by the Ministry of National Education (Erten, 2015; Karaca, 2014), leaf tests related to "*Let's Solve the Puzzle of Our Body*" unit and the exams conducted by the Ministry of National Education for the all secondary school students across the country.

### 2.3 General Information about Unit

In the context of Let's Solve the Puzzle of Our Body unit there are three subtopics: Nutrients and their characteristics, digestion of nutrients and excretory system in our body. In the Ministry of National Education Science Curriculum in total 13 objective were given as 36 lesson hours. The sub topics and contents related to the unit are shown in Table 2.

Table 2. Content of the "*Let's Solve the Puzzle of Our Body*" unit in science curriculum (MoNE, 2015)

Unit Subjects	Subject/Concepts	Lesson Hours	The Number of Objectives
Nutrients and Their Characteristics	Nutriments, balanced nutrition, harms of smoking and alcohol	12	6
Digestion of Nutrients	Structures and organs of digestion, transportation of nutrients in body, nutrients digestion, tooth and dental health	12	4
Excretory System in Our Body	Structures and organs of excretory, structures and organs that enable removal of effluents and noxious substances out of body, types of excretory, kidney health	12	3

### 2.4 Data Analysis

In the analysis of the data obtained during the development of the test, for each item, standard deviation, arithmetic mean, item distinctiveness, item difficulty, Kolmogorov-Smirnov test for the normality test, biserial correlation coefficients in item total score correlation and KR-20 reliability coefficient in reliability calculations were used and calculated statistically.

### 3. Findings

According to Çelik (2000) there are steps to follow when developing an achievement test; planning, item writing, item analysis and item selection. After investigating the research on the achievement test development in the field (Bakioğlu, Karamustafaoglu, S., & Karamustafaoglu, O., 2014; Çalik & Ayas, 2003; Tosun & Taşkesenligil, 2011) and examining the related literature, the steps taken during the test development process in this study are summarized in Figure 1 briefly.

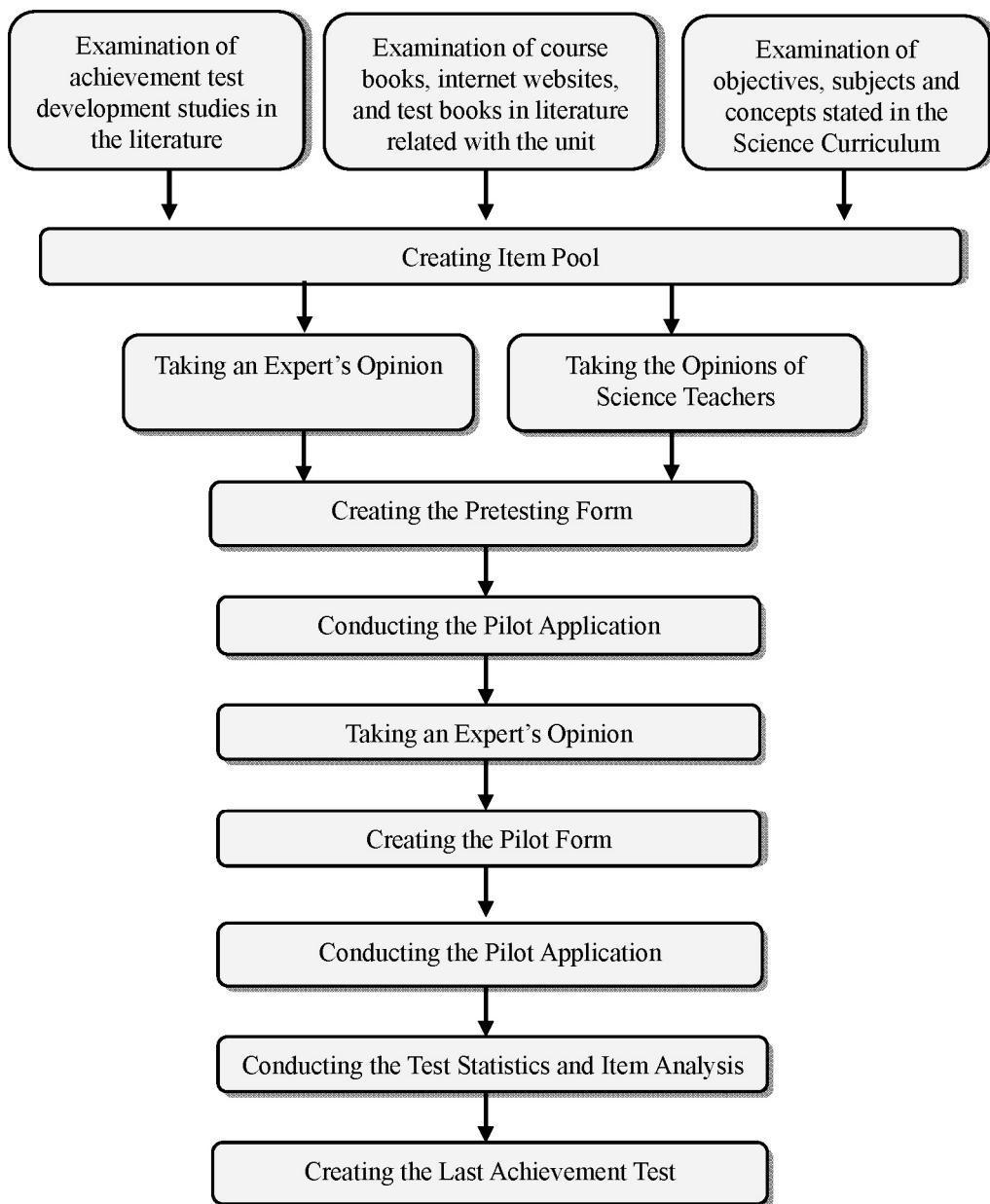


Figure 1. The process of developing achievement test

### 3.1 The Study of Validity

While preparing the achievement test for the “*Let’s Solve the Puzzle of Our Body*” unit in the research, at least three test items related to objective were formed. While creating the items, an expert was consulted in order to ensure validity.

For the content validity of the achievement test two faculty members from Ondokuz Mayıs University Science Teaching Department, four doctoral students and four science teachers, totally 10 people were consulted. For each item found in the pilot achievement test, a graded “*Expert Evaluation Form*” was given to the experts. For each item in this form three grades were given: appropriate, must-be-corrected and must-be-excluded. According to the opinions obtained from the opinion form, the Content Validity Rates (CVR) were calculated for each item (formula 1). CVR is calculated by subtracting one from the division of the number of experts who marked the “*required*” option to the half of number of total experts (Yurdagül, 2005).

$$\text{formula 1} \quad \text{CVR} = \frac{\text{NA}}{\text{N}/2} - 1$$

**NA:** The number of experts who are approving the test items as appropriate.

**N:** The total number of experts who states opinions related to test items.

**CVR:** Content Validity Rates.

In Table 3, minimum values of CVR at  $\alpha=0.05$  significance level are included for an expert opinion according to Veneziano and Hooper (1997). When interpreted according to this table; 10 expert opinions are used in the content validity calculations of the achievement test questions used in this study, therefore, to provide significance statistically according to expert numbers, for 10 experts 0.62 value was used as the Content Validity Criterion (CVC).

Table 3. Minimum values of CVR according to expert opinion number (Veneziano & Hooper, 1997)

Number of Specialist	Min. Value	Number of Specialist	Min. Value	Number of Specialist	Min. Value
5	0.99	10	0.62	15	0.49
6	0.99	11	0.59	16	0.42
7	0.99	12	0.56	17	0.37
8	0.78	13	0.54	18	0.33
9	0.75	14	0.51	19	0.31

In the study conducted, all items from the 46-item achievement test were taken into the application form since no item had a lower value than 0.62 which is the Content Validity Criterion (CVC) for 10 experts. Afterwards, CVRs were collected and the total validity index of the scale was obtained. As a result of the calculations, the Content Validity Index (CVI) of the scale was found to be 0.95 and since  $\text{CVI} \geq \text{CVR}$ , the content validity of the scale was found to be significant statistically (Yurdagül, 2005).

For the face validity of the achievement test, a faculty member from the Department of Science Education, a Science teacher and a language expert were consulted and the necessary corrections were made in the direction of incoming feedbacks. According to these feedbacks, some distractors are at a level that students have difficulty in understanding and some questions are very long and there are two negations in the same question. As a result of the expert examination, no item was eliminated and the pilot application was prepared by making the corrections in the direction of suggestions. So as to determine the content validity of the test, the indicator chart which consists of the unit objectives has been prepared and each one of the objectives has at least three items. The indicator chart related to the content validity has been given Table 4.

The pilot application of a total 46 multiple choice test with 22 items related to “*Nutrients and their Characteristics*”, 13 items of “*Digestion of Nutrients*” and 11 items of “*Excretory System in Our Body*” included in Let’s Solve the Puzzle of Our Body unit was carried out (Table 4).

Table 4. The distribution of test items according to subjects and objectives

Subjects	Objectives	Item No	Number of Item
Nutrients and Their Characteristics	It recognizes that nutrient content is essential for the vital activities of living things.	1, 2, 3, 4, 5, <b>6*</b>	
	It searches and provides information on which nutrients have the most vitamins.	7, 8, <b>9*</b> , <b>10*</b>	
	It deduces that water and minerals are present in all nutrients.	<b>11*</b> , 12, 13	
	It searches and presents the effects of balanced nutrition on human health.	14, 15, 16	22
	It discusses the importance of freshness and naturalness of nutrients for a healthy life based on the research data.	17, 18, <b>19*</b>	
	It discusses the damage of smoking and alcohol to the body based on the research data.	20, 21, 22	
Digestion of Nutrients	It demonstrates the position of structures and organs in digestion on the model respectively.	23, 24, <b>25*</b> , 26, <b>36*</b>	
	It explains the types of teeth by showing them on the model.	27, 28, 29, 31	13
	It cares for nutrition, cleaning and regular teeth control for dental health.	30, 32, 33	
	It deduces that nutrients are transported by blood in body after digestion.	34, 35, 37	
Excretory System in Our Body	It recognizes the structures and organs in excretion.	<b>36*</b> , 37, 38, 43	
	It deduces that there are different types of excretion in the body and that harmful substances emerged as result of the excretory activities must be thrown out of the body.	39, 40, <b>41*</b> , 42, 44	11
	It searches and presents what must be considered to protect kidney health.	43, 45, 46	
			<b>Total</b>
			46

\* Items eliminated as result of pilot application.

### 3.2 Normality Test

Kolmogorov-Smirnov which is one of the normality tests was applied to test the suitability of the normal distribution of the data obtained from the achievement test. The fact that the p value calculated as the result of the analysis is higher than .05 is interpreted as the scores do not show any significant (extreme) deviation from the normal distribution at this significance level (Büyüköztürk, 2010). Accordingly, the Kolmogorov-Smirnov test results show that achievement test scores of the students does not show any significant difference from the normal distribution ( $D(178)=.047$ ;  $p=.200$ ;  $p<.05$ ).

### 3.3 The Item Difficulty and Distinctiveness

In the rating of the results obtained from the achievement test, the total score of each student was calculated by giving 1 point to the correct answers and 0 point to the wrong answers, unanswered questions and to those who marked more than one answer for the same question. The test results obtained after rating are ranked from the highest to the lowest. Item analysis was performed by creating groups in a way that the first 27% ( $N=48$ ) of the score ranking constitute the upper group and the last 27% constitute the lower group and by using Microsoft Excel and SPSS programs for the answers given by the students for each item.

About the levels of item difficulty, it is considered that if the item difficulty index ( $p_j$ ) is between 0.00-0.19 the item is very difficult, if it is between 0.20-0.34 the item is difficult, if it is between 0.35-0.64 the item has medium difficulty, if it is between 0.65-0.79 the item is easy and if it is between 0.80-1.00 the item is very easy (Sözbilir, 2010). In the results of item analysis related to each item in the achievement test, item difficulty index values vary from 0.30 to 0.74.

Item distinctiveness is the comparison of the average of the scores that end groups such as upper and lower groups give each item when they are ranked from the highest to the lowest according to the total scores obtained from the

scale (Tavşancıl, 2006). As a result of the item analysis related to each item in the achievement test, in choosing to decide which item will remain in the test, as item distinctiveness index ( $r_{jx}$ ) it is considered that if  $r_{jx} \leq 0.19$  the item is unacceptable, if it is between 0.20-0.29 the item must be revised, if it is between 0.30-0.39 the item is good/acceptable and if  $0.40 \leq r_{jx}$  the item is very good/acceptable (Özçelik, 2010). In the achievement test developed as 46-items, arithmetic mean, standard deviation, variance, reliability, item distinctiveness, item difficulty, item correlations calculations were performed. Items whose distinctiveness index is lower than 0.30 (9<sup>th</sup>, 10<sup>th</sup>, 11<sup>th</sup>, 19<sup>th</sup>, 25<sup>th</sup>, 36<sup>th</sup>, 41<sup>th</sup>) were excluded from the test. However, 7<sup>th</sup> item whose distinctiveness index is 0.27 was not excluded because four items in total were created (7<sup>th</sup>, 8<sup>th</sup>, 9<sup>th</sup>, 10<sup>th</sup>) related to the objective it qualifies. Other items (9<sup>th</sup>, 10<sup>th</sup>) were eliminated since their distinctiveness were low and 7<sup>th</sup> item was not excluded since in case of its elimination there would have been only one item (8<sup>th</sup>) related to the relevant objective. The distinctiveness of the 6<sup>th</sup> item is 0.31. However, there are 6 items in total related to the objective it qualifies (1<sup>th</sup>, 2<sup>th</sup>, 3<sup>th</sup>, 4<sup>th</sup>, 5<sup>th</sup>, 6<sup>th</sup>). For this reason, the elimination of the 6<sup>th</sup> item, which has the lowest distinctiveness among these items, was deemed appropriate (Table 5).

Table 5. The item analysis of pilot test according to upper-lower group correct answer scores

Item No	Upper group correct answer score	Lower group correct answer score	pj	Difficulty Level	r <sub>jx</sub>	Result	Status
1	46	24	0.73	Easy	0.46	VG	✓
2	45	17	0.65	Easy	0.58	VG	✓
3	46	19	0.68	Easy	0.56	VG	✓
4	40	12	0.54	Average	0.58	VG	✓
5	36	8	0.46	Average	0.58	VG	✓
6	34	19	0.55	Average	0.31	G	-
7	21	8	0.30	Hard	0.27	R	✓
8	33	13	0.48	Average	0.42	VG	✓
9	33	22	0.57	Average	0.23	R	-
10	24	15	0.41	Average	0.19	D	-
11	8	6	0.15	Very Hard	0.04	D	-
12	42	11	0.55	Average	0.65	VG	✓
13	26	9	0.36	Average	0.35	G	✓
14	43	18	0.64	Average	0.52	VG	✓
15	39	17	0.58	Average	0.46	VG	✓
16	38	17	0.57	Average	0.44	VG	✓
17	43	25	0.71	Easy	0.38	G	✓
18	44	20	0.67	Easy	0.50	VG	✓
19	25	14	0.41	Average	0.23	R	-
20	26	4	0.31	Hard	0.46	VG	✓
21	45	16	0.64	Average	0.60	VG	✓
22	35	11	0.48	Average	0.50	VG	✓
23	42	10	0.54	Average	0.67	VG	✓
24	41	20	0.64	Average	0.44	VG	✓
25	20	6	0.27	Hard	0.29	R	-
26	46	19	0.68	Easy	0.56	VG	✓
27	41	21	0.65	Easy	0.42	VG	✓
28	42	11	0.55	Average	0.65	VG	✓

<b>29</b>	44	20	0.67	Easy	0.50	VG	✓
<b>30</b>	47	23	0.73	Easy	0.50	VG	✓
<b>31</b>	35	10	0.47	Average	0.52	VG	✓
<b>32</b>	25	9	0.35	Average	0.33	G	✓
<b>33</b>	47	23	0.73	Easy	0.50	VG	✓
<b>34</b>	38	8	0.48	Average	0.63	VG	✓
<b>35</b>	43	13	0.58	Average	0.63	VG	✓
<b>36</b>	18	8	0.27	Hard	0.21	R	-
<b>37</b>	36	9	0.47	Average	0.56	VG	✓
<b>38</b>	34	13	0.49	Average	0.44	VG	✓
<b>39</b>	30	13	0.45	Average	0.35	G	✓
<b>40</b>	41	14	0.57	Average	0.56	VG	✓
<b>41</b>	23	11	0.35	Average	0.25	R	-
<b>42</b>	34	14	0.50	Average	0.42	VG	✓
<b>43</b>	44	21	0.68	Easy	0.48	VG	✓
<b>44</b>	34	9	0.45	Average	0.52	VG	✓
<b>45</b>	33	17	0.52	Average	0.33	G	✓
<b>46</b>	32	15	0.49	Average	0.35	G	✓
<b>Average</b>			<b>0.52</b>			<b>0.44</b>	

VG: Very good, G: Good, R: to be Revised, D: to be Discarded.

### 3.4 The Item Correlation

In item total correlation, biserial correlation coefficient was used. Biserial correlation coefficient is used to calculate the amount of the relationship between a continuous variable and a variable which is actually continuous but was made discontinuous and artificially with two categories (Büyüköztürk, Çokluk, & Köklü, 2010). In this context, there is a relationship between the score obtained from the sum of the achievement test (continuous variable) and the score obtained from each item of the test. Biserial correlation coefficient was calculated for each item in the test by giving 1 point to the correct answers and 0 point to the wrong and unanswered questions.

Item total correlation explains the relationship between the total score that respondents receive from the assessment instrument and the score they receive from each item. The fact that item total correlation is positive and high indicates that scale items show similar behaviour and that internal consistency of the test is high (Büyüköztürk, 2010). If the total score and correlation of any item is low, it indicates that that item scales a different feature than the other items. Item total correlation should not be negative and it must be at least 0.20. When biserial correlation coefficient of each item included in the achievement test was calculated, the correlation coefficient values of the 9<sup>th</sup>, 10<sup>th</sup>, 11<sup>th</sup>, 19<sup>th</sup> and 41<sup>th</sup> items were found to be below 0.30. These question items are items that were eliminated in the calculation of the item difficulty and distinctiveness made earlier since their values were low. The fact that the correlation between items is high indicates that items are homogeneous and therefore highly reliable (Tavşancıl, 2006). After the item elimination of the achievement test was performed, the distribution of the test items according to the subjects and objectives included in the unit is stated in Table 6.

Table 6. The distribution of test items according to subjects and objectives

Subjects	Objectives	Item No	Number of Item
Nutrients and Their Characteristics	It recognizes that nutrient content is essential for the vital activities of living things.	1, 2, 3, 4, 5	17
	It searches and provides information on which nutrients have the most vitamins.	6, 7	
	It deduces that water and minerals are present in all nutrients.	8, 9	
	It searches and presents the effects of balanced nutrition on human health.	10, 11, 12	
	It discusses the importance of freshness and naturalness of nutrients for a healthy life based on the research data.	13, 14	
Digestion of Nutrients	It discusses the damage of smoking and alcohol to the body based on the research data.	15, 16, 17	13
	It demonstrates the position of structures and organs in digestion on the model respectively.	18, 19, 20	
	It explains the types of teeth by showing them on the model.	21, 22, 23, 25	
	It cares for nutrition, cleaning and regular teeth control for dental health.	24, 26, 27	
	It deduces that nutrients are transported by blood in body after digestion.	28, 29, 30*	
Excretory System in Our Body	It recognizes the structures and organs in excretion.	30*, 31, 35*	9
	It deduces that there are different types of excretion in the body and that harmful substances emerged as result of the excretory activities must be thrown out of the body.	32, 33, 34, 36	
	It searches and presents what must be considered to protect kidney health.	35*, 37, 38	
		<b>Total</b>	38

\* Test items that scales more than one objective.

A total of 38 multiple-choice test items, 17 of which are related to “*Nutrients and their characteristics*”, 13 of which are related to “*Digestion of nutrients*” and 9 of which are related to “*Excretory in Our Body*” are included in the final achievement test (Table 6).

The arithmetic means and standard deviation values of the items of the test finalized according to the item analysis performed as a result of the pilot application of the pilot achievement test are given in Table 7. As a result of the item analysis of the achievement test, 6<sup>th</sup>, 9<sup>th</sup>, 10<sup>th</sup>, 11<sup>th</sup>, 19<sup>th</sup>, 25<sup>th</sup>, 36<sup>th</sup>, 41<sup>th</sup> items were excluded and the difficulty and distinctiveness values stated in Table 7 for each of the other 38 items in the test were received. As a result of the item analysis, the distinctiveness of all questions was calculated above 0.30.

Table 7. The item analyse results of last test

Item No*	Item No**	Upper group correct answer score	Lower group correct answer score	r <sub>jx</sub>	Level of distinctiveness	p <sub>j</sub>	Level of difficulty
1	1	46	25	0.44	VG	0.74	Easy
2	2	45	18	0.56	VG	0.66	Easy
3	3	46	19	0.56	VG	0.68	Easy
4	4	40	9	0.65	VG	0.51	Average
5	5	36	8	0.58	VG	0.46	Average
7	6	22	7	0.31	G	0.30	Hard
8	7	32	15	0.35	G	0.49	Average
12	8	41	11	0.63	VG	0.54	Average
13	9	27	7	0.42	VG	0.35	Hard

<b>14</b>	<b>10</b>	40	20	0.42	VG	0.63	Average
<b>15</b>	<b>11</b>	40	17	0.48	VG	0.59	Average
<b>16</b>	<b>12</b>	38	16	0.46	VG	0.56	Average
<b>17</b>	<b>13</b>	44	26	0.38	G	0.73	Easy
<b>18</b>	<b>14</b>	44	20	0.50	VG	0.67	Easy
<b>20</b>	<b>15</b>	26	5	0.44	VG	0.32	Hard
<b>21</b>	<b>16</b>	44	17	0.56	VG	0.64	Average
<b>22</b>	<b>17</b>	34	11	0.48	VG	0.47	Average
<b>23</b>	<b>18</b>	42	11	0.65	VG	0.55	Average
<b>24</b>	<b>19</b>	43	21	0.46	VG	0.67	Easy
<b>26</b>	<b>20</b>	46	21	0.52	VG	0.70	Easy
<b>27</b>	<b>21</b>	44	20	0.50	VG	0.67	Easy
<b>28</b>	<b>22</b>	44	10	0.71	VG	0.56	Average
<b>29</b>	<b>23</b>	45	19	0.54	VG	0.67	Easy
<b>30</b>	<b>24</b>	46	23	0.48	VG	0.72	Easy
<b>31</b>	<b>25</b>	34	10	0.50	VG	0.46	Average
<b>32</b>	<b>26</b>	26	10	0.33	G	0.38	Average
<b>33</b>	<b>27</b>	47	24	0.48	VG	0.74	Easy
<b>34</b>	<b>28</b>	39	7	0.67	VG	0.48	Average
<b>35</b>	<b>29</b>	43	14	0.60	VG	0.59	Average
<b>37</b>	<b>30</b>	36	6	0.63	VG	0.44	Average
<b>38</b>	<b>31</b>	33	14	0.40	VG	0.49	Average
<b>39</b>	<b>32</b>	32	13	0.40	VG	0.47	Average
<b>40</b>	<b>33</b>	41	14	0.56	VG	0.57	Average
<b>42</b>	<b>34</b>	35	15	0.42	VG	0.52	Average
<b>43</b>	<b>35</b>	43	22	0.44	VG	0.68	Easy
<b>44</b>	<b>36</b>	35	7	0.58	VG	0.44	Average
<b>45</b>	<b>37</b>	34	15	0.40	VG	0.51	Average
<b>46</b>	<b>38</b>	32	16	0.33	G	0.50	Average
				<b>0.49</b>	VG	<b>0.56</b>	<b>Average</b>

\* The item numbers of pilot test.

\*\* The item numbers of last test

VG: Very good, G: Good.

Table 8. Achievement test statistics found as a result of item analysis

Achievement Test	Number of Item	N	Mean	Variance	Std. Deviation	Average Difficulty	KR-20
Pilot Test	46	178	23.83	66.65	8.16	0.52	0.86
Last Test	38	178	21.08	57.81	7.60	0.56	0.87

Arithmetic mean, standard deviation, variance, difficulty and reliability calculations of 38 items were repeated in the final achievement test (Table 8). In the research, the average difficulty of the pilot and final achievement tests was found to be moderate. The average difficulty of the achievement tests must be 0.50 so that they can serve the feature that is scaled and they can be highly reliable (Kan, 2012).

### 3.5 The Study of Reliability

In the reliability calculation of the achievement test, the KR-20 reliability coefficient was calculated. The KR-20 is suitable for determining the reliability coefficient of tests in which each item is parallel to each other, which has the same mean and variance and which was scored by giving one point to the correct answers for each question, and not giving any point to the wrong answers or unanswered questions (Baykul, 2010; Tekin, 2000). The reliability coefficient value was calculated as 0.86 as a result of the Kuder Richardson 20 (KR-20) calculation of the pilot achievement test whereas, after the elimination of the eight items as result of the item analysis KR-20 reliability coefficient was calculated as 0.87 (Table 8). An assessment instrument whose KR-20 reliability coefficient is 0.70 or higher is acknowledged as reliable (Fraenkel & Wallen, 2006; Özçelik, 2010; Saipanish, Hiranyatheb, & Lotrakul, 2015). Therefore, this achievement test is considered as reliable. As a result of the item analysis, the achievement test consisting of 38 multiple choice items was finalized and prepared for using in the research.

## 4. Conclusion and Suggestions

The assessment and evaluation process is important in terms of assessing the effectiveness of science teaching. One of the frequently used assessment instruments in the assessment and evaluation studies is multiple-choice achievement tests. Today multiple choice tests are one of the most widely used assessment instruments which allows comprehensive assessment of achievement and easy scoring for the practitioner by providing many questions in a short period of time (Burton, Sudweeks, Merrill, & Wood, 1991; Bağcan Büyükturan & Çikrikçi Demirtaşlı, 2012; Treagust, 1988). Test is an assessment instrument easy to apply and score in the assessment and evaluation process since it consists of multiple choice items. For this reason, the aim of this study is to develop a reliable and valid assessment instrument which can assess the achievement of students related to the fifth grade science course "*Let's Solve the Puzzle of Our Body*" unit.

In the process of developing the test, firstly the pilot application of the test and test and item analysis were performed. As a result of the item analysis of the achievement test consisting of 46 items in total, the final test consisting of 38 items was created by eliminating 8 items. A table of specifications showing the relationship between the test items created in terms of content validity and the objectives included in the Ministry of National Education Science Curriculum was prepared. In addition, the Content Validity Index (CVI) of the test was calculated to be 0.95 by taking expert opinions for each item. As a result of the item analyses carried out during the test development process; item difficulties were calculated between 0.30-0.74, item distinctiveness index were calculated between 0.31-0.71, and item-total score biserial correlation coefficients were calculated between 0.30-0.66. While calculating the KR-2 reliability coefficient of the final test, the average difficulty of the test was found to be moderate and its average distinctiveness was found to be very good. The results show that the achievement test is reliable and valid in terms of evaluating the academic achievements of the fifth grade students related to the "*Let's Solve the Puzzle of Our Body*" unit.

When the literature on digestive system, excretory system, nutrients and nutrient types are examined, it is seen that college students carried out studies on 6<sup>th</sup> and 7<sup>th</sup> grade students at secondary school (Alkan Dilbaz, 2013; Güçlüer, 2012; Güngör & Özgür, 2009; Patrick & Tunnicliffe, 2010; Prokop & Faněovičová, 2006; Yıldırım, 2012). This test will enable Piaget to identify the deficiencies in knowledge in the biology field during the transitional period of the 5<sup>th</sup> grade students transitioning from the concrete process period to the abstract process period.

It is believed that this assessment instrument can help to determine the readiness level of the 5<sup>th</sup> grade students and their lack of knowledge in subtopics and that it can help the scientific studies of the researchers conducting experimental research. In the direction of the results obtained from this research, the following suggestions have been made:

- With this developed achievement test, level and lack of knowledge of the students in "*Nutrients and Their Characteristics*", "*Digestion of Nutrients*" and "*Excretory System in Our Body*" subtopics included in "*Let's Solve the Puzzle of Our Body*" unit can be determined in the transition period of the students to secondary school.
- The developed achievement test can help students to organize their learning activities according to their determined deficiencies by determining their readiness and deficiencies in terms of the 5<sup>th</sup> grade biology subjects.
- With the developed achievement test, it is possible to determine the misconceptions in the students by examining their level of knowledge and deficiencies as well as the questions they answered wrong. Because

the distracters of each item in the test were prepared according to the misconceptions that students have in relation to the topic.

- The developed achievement test can be used as a data collection tool for other researches to be carried out in the field of science education.

### Acknowledgments

This research was produced from first author's PhD thesis and supported by Scientific Research Project Number PYO.EGF.1904.13.011 within the University Project of Supporting Program for Master Theses in Ondokuz Mayıs University.

### References

Akgündüz, D. (2013). *The effect of blended learning and social media-supported learning on the students' academic achievement, motivation, attitude and self-learning skill in science education* (PhD Thesis). Educational Sciences Institute, Science Education Program, Marmara University, İstanbul.

Alkan Dilbaz, G. (2013). *The effects of inquiry-based learning on attitude, academic success, problem solving and inquiry skills* (Master Thesis). Educational Sciences Institute, Division of Curriculum Instruction, Mersin University, Mersin.

Aydede, M. N., & Matyar, F. (2009). Aktif öğrenme yaklaşımının fen bilgisi dersindeki akademik başarı ve kalıcılığa etkisi. *Kastamonu Journal of Education*, 17(1), 137-152.

Bağcan Büyükturan, E., & Çıraklı Demirtaşlı, S. (2012). Comparing the psychometric characteristics of multiple choice tests and structural communication grids. *Journal of Faculty of Educational Sciences*, 45(1), 395-415.

Bakioğlu, B., Karamustafaoglu, S., & Karamustafaoglu, O. (2014). *5<sup>th</sup> Class "our body: Let's solve the puzzle" unit achievement test: Validity and reliability* (p. 271). ICEMST 2014.

Banet, E., & Nunez, F. (1997). Teaching and learning about human nutrition: A constructivist approach. *International Journal of Science Education*, 19(10), 1169-1194. <https://doi.org/10.1080/0950069970191005>

Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması* (2nd ed.). Ankara: PegemA Publishing.

Burton, S. J., Sudweeks, R. R., Merrill, P. F., & Wood, B. (1991). *How to prepare better multiple-choice test items: Guidelines for university faculty*. Brighaam young university testing.

Büyüköztürk, Ş. (2010). *Sosyal bilimler için veri analizi el kitabı* (12th ed.). Ankara: Pegem Akademi.

Büyüköztürk, Ş., Çokluk, O., & Köklü, N. (2010). *Sosyal bilimler için istatistik* (6th ed.). Ankara: Pegem Akademi Publishing.

Çalık, M., & Ayas, A. (2003). Çözeltilerde kavram başarı testi hazırlama ve uygulama. *Pamukkale University Journal of Education*, 14(2), 1-17.

Çardak, O. (2005). Student science teachers' ideas of the digestive system. *Journal of Education and Training Studies*, 3(5), 127-133.

Carvalho, G. S., Silva, R., & Clément, P. (2003). *Epistemological and didactical learning obstacles identified in Portuguese primary school pupils (Synopsis)*, *Research and the Quality of Science Education*. ESERA, Noordwijkerhout, CD.

Çelik, D. (2000). *Okullarda ölçme ve değerlendirme nasıl olmalı* (pp. 50-86)? İstanbul: Milli Eğitim Basımevi.

Erdoğan, S. (2010). *The effect of the educational drama method on students? Achievement in? The systems of body structures? Unit in science and technology lesson* (Master Thesis). Educational Sciences Institute, Gazi University, Ankara.

Erten, F. (2015). *Ortaokul 5. sınıf fen bilimleri ders kitabı*. Ankara: Özgün Matbaacılık Sanayi ve Ticaret AŞ.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education*. New York: McGraw-Hill.

Griffard, P. B. (2001). The Two-tier instrument on photosynthesis: What does it diagnose? *International Journal of Science Education*, 23(20), 1039-1052. <https://doi.org/10.1080/09500690110038549>

Güçlüer, E. (2012). *The effect on success, attitude and scientific process skills of the use of scientific literacy developing activities in the unit named “systems of our body” in science and technology course*. Educational Sciences Institute, Science Education Program, Dokuz Eylül University, İzmir.

Güngör, B. (2009). *A longitudinal study with seventh class students to identify the origins of the misconceptions about human digestive system* (PhD Thesis). Sciences Institute, Balıkesir University, Balıkesir.

Güngör, B., & Özgür, S. (2009). The Causes of the fifth grade students' misconceptions originated from didactic about digestive system. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, 3(2), 149-177.

İnel, D. (2009). *The effects of the using of problem based learning method in science and technology course on students? The levels of constructing concepts, academic achievements and enquiry learning skill perceptions* (PhD Thesis). Educational Sciences Institute, Dokuz Eylül University, İzmir.

Kan, A. Ü. (2012). *The effects of using individual and group mind mapping on students' academic achievement, retention and affective characteristics in social studies course* (Unpublished PhD Thesis). Educational Science Institute, Fırat University, Elazığ.

Karaca, Ö. (2014). *Ortaokul fen bilimleri 5. sınıf ders kitabı*. Ankara: KOZA Yayın Dağıtım AŞ.

Kiras, B. (2013). *Effect of active learning methods on student's achievement, attitude and creativity* (Master Thesis). Sciences Institute, İstanbul University, İstanbul.

Marx, R. W. et al. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching*, 41(10), 1063-1080. <https://doi.org/10.1002/tea.20039>

Ministry of National Education (MoNE). (2015). *İlköğretim kurumları (ilkokullar ve ortaokullar) fen bilimleri dersi (3,4,5,6,7 ve 8. sınıflar) öğretim programı*. Ankara: MEB Talim ve Terbiye Kurulu Başkanlığı.

Nunez, F., & Banet, E. (1997). Students' conceptual patterns of human nutrition. *International Journal of Science Education*, 19(5), 509-526. <https://doi.org/10.1080/0950069970190502>

Özçelik, D. A. (2010). *Test hazırlama kılavuzu* (4th ed.). Ankara: PegemA Publishing.

Patrick, P. G., & Tunnicliffe, S. D. (2010). Science teachers' drawings of what is inside the human body. *Journal of Biological Education*, 44(2), 81-87. <https://doi.org/10.1080/00219266.2010.9656198>

Prokop, P., & Faněovičová, J. (2006). Student's ideas about the human body: Do they really draw what they know? *Journal of Baltic Science Education*, 10.

Saipanish, R., Hiranyatheb, T., & Lotrakul, M. (2015). Reliability and validity of the thai version of the Florida obsessive-compulsive inventory. *The Scientific World Journal*, 1-7. <https://doi.org/10.1155/2015/240787>

Sözbilir, M. (2010). *Madde analizi ve test geliştirme*. Retrieved June 6, 2015, from <https://olcmevedegelelendirme.files.wordpress.com/2010/09/7-madde-analizi-ve-test-gelistirme.pdf>

Tavşancıl, E. (2006). *Tutumların ölçülmesi ve SPSS ile veri analizi* (3th ed.). Ankara: Nobel Publishing.

Tekin, H. (2000). *Eğitimde ölçme ve değerlendirme* (14th ed.). Ankara: Yargı Publishing.

Tosun, C., & Taşkesenligil, Y. (2011). Revize edilmiş bloom'un taksonomisine göre çözeltiler ve fiziksel özellikler konusunda başarı testinin geliştirilmesi: Geçerlik ve güvenirlilik çalışması. *Journal of Kastamonu Education*, 19(2), 499-522.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159-169. <https://doi.org/10.1080/0950069880100204>

Veneziano, L., & Hooper, J. (1997). A Method for Quantifying Content Validity of Health-Related Questionnaires. *American Journal of Health Behaviour*, 21(1), 67-70.

Yıldırım, C. (2012). *The effect of scientific process skills activities on elementary school 7th grade student's reflective thinking* (Master Thesis). Science Institute, Science Education Program, Pamukkale University, Pamukkale.

Yurdagül, H. (2005). *Ölçek geliştirme çalışmalarında kapsam geçerliliği için kapsam geçerlilik indeksinin kullanılması*. Ulusal Eğitim Bilimleri Kongresinde Sunulmuş Bildiri, Denizli.

## Appendix A

### The Academic Achievement Test of Let's Solve the Our Body Puzzle Unit

#### Question 1)

Which of these nutrition below are the most **fuel nutrients** in comparison to the others?

- A. Rice-Sugar
- B. Cracked wheat-Orange
- C. Meat-Egg
- D. Pasta-Spinach

#### Question 2)

*"Selin had cheese, bread and honey on breakfast; pasta on her lunch; French fries and rice on dinner."*

Regarding the nutriment that Selin ate for a day, try to find out what kind of food Selin takes in her body **excessively**?

- A. Carbohydrate
- B. Protein
- C. Fat
- D. Vitamin

#### Question 3)

Which nutriment group is **most important as regulator** in our body?

- A. Protein
- B. Fat
- C. Carbohydrate
- D. Water

#### Question 4)

What kinds of nutrient do we get from the energy that our body needs **primarily** to think, talk, walk, play sports, and so on?

- A. Protein
- B. Vitamin
- C. Carbohydrate
- D. Fat

#### Question 5)

Think about the nutrient groups that found excessively on animal nutrient such as meat, milk, egg, fish and cheese.

Which of these below is **not one of the primary duties** of the nutrient group you think about?

- A. It provides growth and development
- B. It has an important role in the development of intelligence
- C. It has a serious role in the defence against germs
- D. They provide the energy that body

#### Question 6)

*"Citrus, strawberry, tomato, parsley, cabbage, rosehip"*

Think about the type of the vitamin found in these kinds of nutrients above. Which one of the following disorders may come in **the lack of this type of vitamin**?

- A. Increase in teeth and gum problems
- B. Haemophilia
- C. Anaemia, fatigue, scars on the skin
- D. Liver, cardiovascular diseases

#### Question 7)

If you think about the types of vitamins contained in the following food groups, which of the given nutrient groups contains the **Vitamin K more**?

- A. Citrus, tomato, strawberry
- B. Red meat, green vegetables, banana and peach
- C. Carrot, wheat, legume family and peanut
- D. Fish, dairy products

#### Question 8)

Which of the following is the group of nutrient that need to be taken every day for a healthy body and that are found in all nutrients as regulatory?

- A. Protein-Minerals
- B. Water-Minerals
- C. Carbohydrate-Water
- D. Protein-Water

**Question 9)**

- I.** Used for energy
- II.** They are not synthesized in our bodies but taken from the outside
- III.** They exist in all the food we eat
- IV.** They work as constructive and reparative.

Which of the information above about water and minerals are true?

- A.** I.-III.- IV.
- B.** II.-IV.
- C.** II.-III.
- D.** III.-IV.

**Question 10)**

The doctor asks Betül, who is sick, what she eats.

**Betül:**

-“*I often eat hamburger and drink coke.*”

**Doctor:**

-“*If you keep eating like that, you are going to get a skin breakdown and feel fatigue*”

What is the doctor trying to say Betül **basically**?

- A.** She should have a vitamin-based diet instead of carbohydrate
- B.** She should have a protein-based diet
- C.** She consumes too much vitamin
- D.** To eat one type of food is unhealthy.

**Question 11)**

Which of these below does not belong to a person who eats properly?

- A.** He consumes less food and loses weight
- B.** Resistant to diseases
- C.** Body cells and tissues renew themselves
- D.** Heart-attack risk is unlikely

**Question 12)**

“*Cenk has an every-day diet based on carbohydrate.*”

If he keeps eating like that, which one is the most likely to happen to his body?

- A.** Excess carbohydrate improves his body muscles
- B.** Excess carbohydrate is stored as vitamin
- C.** Excess carbohydrate turns into fat and make him gain weight
- D.** As excess carbohydrate is stored, body lacks energy

**Question 13)**

When Kayra goes to the supermarket for shopping which of the following is an **inappropriate** behaviour for a healthy diet?

- A.** Checking the TSI (Turkish Standard Institution) logo on the product
- B.** Choosing the products with additives
- C.** Preferring natural food to frozen food
- D.** Checking the best before dates

**Question 14)**

Which of the things below is **not suitable** for a balanced nutrition?

- A.** We should only eat fruits and vegetables.
- B.** We should drink plenty of water
- C.** We should consume different nutrients on every meal

We should consume nutrients in proper amount according to the age and physical activity

**Question 15)**

Which of the things below is **not** one of the harms of smoking?

- A.** It causes lung and laryngeal cancer
- B.** Cardiovascular diseases
- C.** Causes difficulty in talking and slow reflexes
- D.** Causes to skin breakdown

**Question 16)**

Which of the statements below is **not** one of the harms of alcohol?

- A. It weakens the will by negatively affecting the nervous system
- B. It affects the brain, muscles and veins adversely
- C. Makes you sleepy and brings an order to sleeping pattern
- D. It makes it hard to control the behaviours and senses

**Question 17)**

Which of the following is the body and structure that alcohol affects the most, negatively?

- A. Lung-Circulatory
- B. Heart- Vein
- C. Liver-Nerve
- D. Kidney-Urinary

**Question 18)**

Which one of the following is the route of food in the digestive system after stomach?

- A. Kidney-Small intestine-Large intestine
- B. Small intestine-Large intestine
- C. Kidney-Small intestine-Large intestine-urinary incontinence
- D. Small intestine-Large intestine-Urinary incontinence

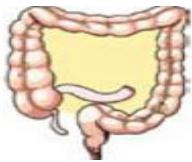
**Question 19)**

Which of the following are physically disintegrating foods in the digestive system in the human body and the remaining waste after digestion is thrown out?

- A. Pharynx-Stomach
- B. Stomach-Large intestine
- C. Pharynx-Large intestine
- D. Mouth-Anus

**Question 20)**

I.



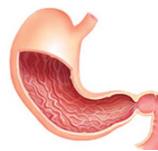
II.



III.



IV.



Which of the organs given above are the organs in charge of digestion?

- A. I-II
- B. II-III
- C. I-III
- D. I-IV

**Question 21)**

Which of the following is the largest number of teeth in an adult individual who can crush and grind food?

- A. Molar tooth
- B. Incisor tooth
- C. Wisdom tooth
- D. Dog tooth

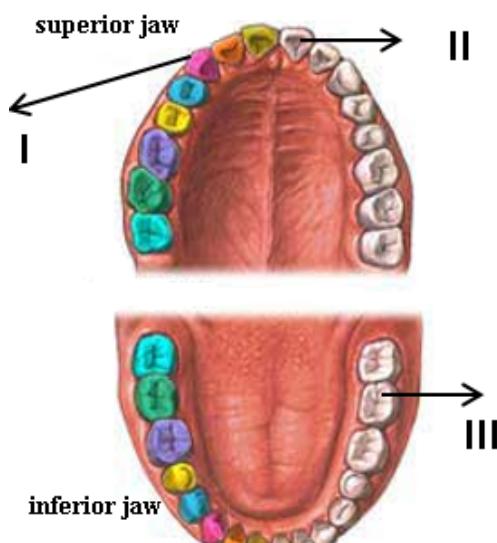
**Question 22)**

Koray used the dough to make a tooth model for the assignment given by the teacher. For this, 4 blue, 8 yellow and 16 red coloured game hurries are used.

Which teeth do the colours in Koray's tooth model represent?

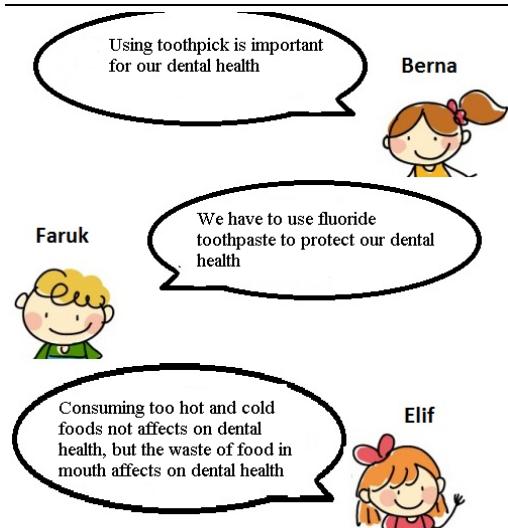
Blue      Yellow      Red

A. Incisor	Dog	Molar
B. Molar	Dog	Incisor
C. Dog	Incisor	Molar
D. Dog	Molar	Incisor

**Question 23)**

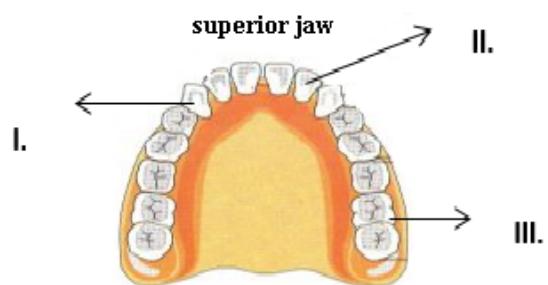
Which kinds of teeth below are shown with arrows in the mouth model above?

I	II	III
A. Dog tooth	Molar tooth	Incisor tooth
B. Incisor tooth	Dog tooth	Molar tooth
C. Dog tooth	Incisor tooth	Molar tooth
D. Incisor tooth	Molar tooth	Dog tooth

**Question 24)**

Which of the above-mentioned statements of Elif can be shown as an example of correct behaviour to protect mouth and dental health?

- A. Berna
- B. Faruk
- C. Elif
- D. All of them

**Question 25)**

In the tooth model given above, which of the following duties of the dental types below indicated by numbers are correctly given?

A.	I.	II.	III.
B.	II.	III.	I.
C.	III.	II.	I.
D.	II.	I.	III.

**Question 26)**

Which of the following statements about decayed teeth is wrong?

- A. Decayed tooth damages the heart.
- B. Germs on decaying tooth cause diseases in internal organs.
- C. Newly decaying teeth should be removed.
- D. Decayed tooth causes foul breath in mouth

**Question 27)**

Which of the following statements the doctor advises to Berk who has toothache is **not correct**?

- A. You should consume less milk and dairy products
- B. You should use fluoride toothpaste
- C. You should consume fresh fruit
- D. Brush your teeth three times a day

**Question 28)**

Which of the following is the part of the body by which food that is digested in our bodies, water, vitamins and minerals absorbed into the circulation system?

- A. Kidney
- B. Small intestine
- C. Stomach
- D. Gullet

**Question 29)**

After the digested food becomes shattered and absorbed, in which way are the beneficial parts carried on the body?

- A. Passes to stomach to get reabsorbed
- B. It spreads throughout the body through the large intestine.
- C. It is transported through the liver to the body
- D. It spreads to the whole body with blood circulation

**Question 30)**

Which one of the following is the system which helps the nutrients mix with blood and the organ where water and mineral are absorbed?

- A. Circulation - Small intestine
- B. Urinary - large intestine
- C. Urinary - Kidney
- D. Digestion - Large intestine

**Question 31)**

Which of the following is the ureter's duty?

- A. The short pipe that the urine is thrown out
- B. Place where the blood is filtered
- C. The place where the urine is collected
- D. The conduit carrying the urine from the kidneys to the urine

**Question 32)**

Which of the following does not play a major role in disposing waste and residual substances in the body?

- A. Sweating
- B. Digestion
- C. Breathing
- D. Urine formation

**Question 33)**

Which of the following is not an organ that helps to remove waste from your body?

- A. Stomach
- B. Lung
- C. Skin
- D. Kidney

**Question 34)**

*“Urea - Oxygen - Sweat - Carbon Dioxide - Urine”*

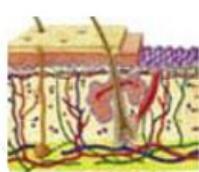
How many of the above are waste materials that are formed in the human body?

- A. 1
- B. 2
- C. 3
- D. 4

**Question 35)**

Which of the following should **not** be done for the health of the drainage system?

- A. We should drink plenty of water.
- B. We must wash our hands with soap after toilet.
- C. We have to bathe frequently.
- D. When we have toilet, we should keep our urine.

**Question 36)****I.****II.****III.****IV.**

Which of the above organs and structures are **urinary system** organs?

- A. III-IV
- B. I-II-IV
- C. II-III
- D. I-II-III

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

**Question 37)**

*"The doctor has some suggestions for Merve's father who suffers from kidney failure and he says he should pay attention to his life."*

Which one of the following is **harmful**?

- A. Eat his meal too much salty
- B. Avoiding cold and especially getting cold feet
- C. Drink plenty of water
- D. Avoid doing sports

**Question 38)**

Kaan will prepare a poster to protect the kidneys' health. Which of the following **cannot** be one of the items that should be mentioned on the poster?

- A. Treatment of tooth decay
- B. Eat bitter and spicy foods
- C. Clean water and cleaned food
- D. Cold protection of urinary tracts